

TARGET DENSITY NORMALIZATION FOR MARKOV CHAIN MONTE CARLO ALGORITHMS

ALLEN CALDWELL, CHANG LIU
MAX PLANCK INSTITUTE FOR PHYSICS, MUNICH, GERMANY

ABSTRACT. Techniques for evaluating the normalization integral of the target density for Markov Chain Monte Carlo algorithms are described and tested numerically. It is assumed that the Markov Chain algorithm has converged to the target distribution and produced a set of samples from the density. These are used to evaluate sample mean, harmonic mean and Laplace algorithms for the calculation of the integral of the target density. A clear preference for the sample mean algorithm applied to a reduced support region is found, and guidelines are given for implementation.

1. INTRODUCTION

Markov Chain Monte Carlo (MCMC) algorithms [1] are often used to generate samples distributed according to non-trivial densities in high dimensional spaces. Many algorithms have been developed that allow MCMCs to produce samples Λ from an unnormalized target density $f(\lambda)$:

$$\Lambda \sim f(\lambda) .$$

In many applications, it is desirable or even necessary to be able to normalize the target density. I.e., to calculate

$$(1) \quad I \equiv \int_{\Omega} f(\lambda) d\lambda$$

where $\Omega \in \mathbb{R}^D$ is the support of f . This integral can be computationally very costly or impossible to perform with standard techniques if the volume where the target f is non-negligible occupies a very small part of the total volume of Ω .

An important area where such integration is necessary is for Bayesian data analysis [2, 3]. Bayes' formula reads, for a given model M ,

$$(2) \quad P(\lambda|\text{Data}, M) = \frac{P(\text{Data}|\lambda, M)P_0(\lambda|M)}{P(\text{Data}|M)}$$

where here λ are the parameters of the model and the data are used to extract probabilities for possible values of λ . The denominator is usually expanded using the Law of Total

Date: October 27, 2014.

Probability and written in the form

$$(3) \quad Z = P(\text{Data}|M) = \int P(\text{Data}|\lambda, M)P_0(\lambda|M)d\lambda$$

and goes by the names ‘evidence’, or ‘marginal likelihood’, and is the type of integral that we want to be able to calculate (here the data are fixed and $f(\lambda) = P(\text{Data}|\lambda, M)P_0(\lambda|M)$). An example use of Z is for the calculation of Bayes Factors in the comparison of two models:

$$\text{BF} \equiv \frac{P(\text{Data}|M_A)}{P(\text{Data}|M_B)} = \frac{Z_A}{Z_B}.$$

Another application where the calculation of a normalization can be very important is in the parallelization of the MCMC algorithm [4]. While the MCMC approach has very attractive features, it is often slow in its execution due to the nature of the algorithm. A goal is therefore to parallelize the computations needed to map out the target density. This looks at first sight difficult since the MCMC algorithms are by construction serial. A parallelization of the calculations can however be achieved via a partitioning of the support. I.e., we partition Ω into sub volumes ω_i with

$$\cup \omega_i = \Omega \quad \omega_i \cap \omega_j = \emptyset \text{ for } i \neq j$$

and we run a separate MCMC sampling for each sub volume ω_i . In order to have a final set of samples representing the target density over the full support, we need to know the relative probabilities for the different sub volumes. I.e., we need

$$I_i \equiv \int_{\omega_i} f(\lambda)d\lambda \quad \left(\sum_i I_i = I \right)$$

The samples in the different regions are then given weights $\propto \frac{I_i N}{I N_i}$ with N_i the number of samples from $f(\lambda)$ in ω_i and $\sum_i N_i \equiv N$.

2. TECHNIQUES

A variety of techniques to calculate the evidence in Bayesian Calculations have been successfully developed. A summary can be found in [5], where a number of MCMC related techniques are reviewed, including Laplace’s method [6], harmonic mean estimation [7], Chib’s method [8], annealed importance sampling techniques [9], Nested Sampling [10] and thermodynamic integration methods [11, 12].

We are here specifically interested in testing techniques directly applicable in an MCMC setting, and which is independent of the specific MCMC algorithm. We assume that the MCMC algorithm has been successfully run to extract samples according to the target density, and the goal is to provide an algorithm for calculating the normalization (or evidence). Given our requirements, only arithmetic mean estimation (AME), harmonic mean estimation (HME) and Laplace methods are directly applicable. Using AME and HME methods directly is known to fail in many situations, and the Laplace method is only applicable if the target density is Gaussian. We introduce the use of a reduced integration volume and normalization using the MCMC output to improve the AME and HME performance.

After a description of the techniques, we report on numerical investigations of the different approaches using samples from the MCMC code BAT [13].

2.1. Reduced Volume Evaluation. Assuming the MCMC has been successfully run to extract samples according to $f(\lambda)$, one of the quantities directly retrievable from the MCMC output is an estimate of the parameter values at the global mode: Λ^* is in the neighborhood of λ^* . I.e., we know approximately where the integrand in Eq. 1 has its maximum.

We note that

$$(4) \quad r \equiv \frac{\int_{\omega} f(\lambda) d\lambda}{I} \approx \frac{N_{\Delta}}{N_{\text{MCMC}}} \equiv \hat{r}$$

with ω a sub support of Ω is directly estimated from the MCMC output by counting the fraction of samples falling within ω , N_{Δ} (the reason for this notation will become clear below). I.e., the task of evaluating I reduces to integrating the function $f(\lambda)$ over a well-chosen region - presumably a small region around λ^* and dividing by \hat{r} . This integral can be much simpler to evaluate than the integral over the full support.

2.2. Choice of Region ω . In the following, we use a simple hypercube for our integration region. From the MCMC samples, we can construct the marginalized distributions along each of the λ dimensions. We define an interval along each dimension centered at Λ^* with width which is a multiple of the standard deviation (we use the symbol Δ to represent this factor). The optimum value of Δ depends on the dimensionality of the problem as described below. Another option would be to produce a covariance matrix of the Λ for sampling using a multivariate normal distribution if desired, but this was not found necessary in the examples we have studied.

2.3. Arithmetic Mean Estimation. The integral in the numerator in Eq. 4 can presumably be determined in a straightforward way since now we are focusing on a small volume with significant mass. The standard importance sampling approximation is given by

$$\begin{aligned} I_{\Delta} &\equiv \int_{\omega} f(\lambda) d\lambda \\ &= \int_{\omega} \frac{f(\lambda)}{g(\lambda)} g(\lambda) d\lambda \\ &\approx \frac{1}{N_{\text{SM}}} \sum_{\Lambda \in \omega} \frac{f(\Lambda)}{g(\Lambda)} \Big|_{\Lambda \sim g(\lambda)} \end{aligned}$$

where our sampling probability density is given by $g(\lambda)$. N_{SM} is the number of samples used in the calculation. If we choose for $g(\lambda)$ a uniform distribution in the hypercube, then we have the well-known sample mean result

$$(5) \quad I_{\Delta} \approx \frac{V_{\Delta}}{N_{\text{SM}}} \sum_{\Lambda \in \omega} f(\Lambda) \equiv \hat{I}_{\Delta}$$

with V_Δ the volume of the hypercube. Our estimator for I is then

$$(6) \quad \hat{I}_{AME} \equiv \frac{\hat{I}_\Delta}{\hat{r}} .$$

We will use this simplest version of the estimator for our examples below.

2.4. Uncertainty Estimate. Assuming unbiased Gaussian distributions for \hat{I}_Δ and \hat{r} about their true values, we can estimate the uncertainty for I with

$$(7) \quad \sigma_I = \sqrt{\left(\frac{\sigma_r \cdot \hat{I}_\Delta}{\hat{r}^2}\right)^2 + \left(\frac{\sigma_{I_\Delta}}{\hat{r}}\right)^2}$$

where

$$\sigma_r = \sqrt{\frac{\hat{r}(1 - \hat{r})}{N_{\text{ESS}}}} .$$

The effective sample size [14] is defined here as

$$(8) \quad N_{\text{ESS}} \equiv \frac{N_{\text{MCMC}}}{1 + 2 \sum_{j=1}^{N_{\text{MCMC}}} \rho_j}$$

with the autocorrelation function at j defined for our MCMC sample as

$$(9) \quad \rho_j \equiv \frac{\sum_{i=1}^D (\Lambda_i^j \Lambda_i^{j+1} - \langle \Lambda_i \rangle^2)}{\sum_{i=1}^D \sigma_i^2}$$

$$(10) \quad \langle \Lambda_i \rangle = \frac{1}{N_{\text{MCMC}}} \sum_{j=1}^{N_{\text{MCMC}}} \Lambda_i^j$$

$$(11) \quad \sigma_i^2 = \frac{1}{N_{\text{MCMC}}} \sum_{j=1}^{N_{\text{MCMC}}} (\Lambda_i^j)^2 - \langle \Lambda_i \rangle^2 .$$

In these equations, the subscript $i = 1 \dots D$ labels the component of Λ , while the index j labels the iteration in the MCMC.

The uncertainty from the sample mean integration is estimated by separating the sample mean calculation of \hat{I}_Δ into K batches and looking at the variance of these calculations:

$$(12) \quad \sigma_{I_\Delta} = \sqrt{\frac{\sum_{k=1}^K \left(\hat{I}_{\Delta,k} - \langle \hat{I}_\Delta \rangle \right)^2}{K(K-1)}} .$$

With these definitions, we are able to report both an estimate for our integral and an uncertainty. These will be compared to accurately calculated values for the chosen examples in the following sections.

2.5. Harmonic Mean Estimation. The HME [7] value for I can be calculated as follows:

$$(13) \quad E \left[\frac{1}{f(\lambda)} \right]_{\hat{f}(\lambda)} = \int_{\Omega} \frac{1}{f(\lambda)} \cdot \frac{f(\lambda)}{I} d\lambda$$

$$(14) \quad = \frac{V}{I}$$

where $\hat{f}(\lambda)$ is the normalised target density and V is the total volume of the support. The HME estimator is then

$$(15) \quad \hat{I} = \frac{N_{\text{MCMC}} V}{\sum_{\Lambda \in \Omega} \frac{1}{f(\Lambda)}}.$$

This calculation is performed directly from the MCMC output from which the samples Λ as well as $f(\Lambda)$ are available, and does not require an extra sample mean calculation as in the AME scheme. However, it can be unstable because of samples occurring (or missing) in regions where $f(\lambda)$ is small (relative to other regions). We can improve the estimation, as originally noted in [15], by limiting ourselves to a small volume around the mode. Using the same notation as above, we can write

$$(16) \quad \hat{I}_{\text{HME}} \equiv \frac{N_{\text{MCMC}} V_{\Delta}}{\sum_{\Lambda \in \omega} \frac{1}{f(\Lambda)}}.$$

where now only the samples in the restricted support ω are used. The uncertainty in the estimate is calculated by separating the MCMC samples included in our integration region into batches and looking at the variation of these estimates.

2.6. Laplace Method. In this approach, the target distribution is assumed to be represented by a (multivariate) Gaussian distribution. The estimator for the normalization is then

$$(17) \quad \hat{I}_{\text{L}} \equiv (2\pi)^{D/2} |\Sigma^*|^{1/2} f(\Lambda^*)$$

where the target density is evaluated at the mode returned from the MCMC and $|\Sigma^*|$ is the determinant of the covariance matrix evaluated numerically from the samples Λ . This method is clearly only expected to work in cases where the assumption of normality is valid.

3. EXAMPLES

3.1. Product of one-dimensional Gaussians. We start with a simple example - the target density is the product of a number of Gauss functions depending on only one parameter - to describe our testing procedures in detail. We then move on to more complicated examples in multivariate spaces, including functions with degenerate modes. All MCMC calculations were performed using the BAT program, with samples from the target density taken after convergence of the MCMC algorithm.

We start with the following target function:

$$(18) \quad f(\mu) = \prod_{i=1}^m \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x_i-\mu}{\sigma}\right)^2}.$$

This type of function could, e.g., be the likelihood function constructed for producing an estimate of a quantity, μ , given m measurements, $\{x_i\}$, with a sampling distribution modeled by a Gaussian probability distribution of fixed width σ . The normalization integral for the target can be performed analytically assuming the volume of interest extends well beyond the extreme values of the x_i . For the more general case of the product of m D -dimensional uncorrelated Gaussian functions with known variances σ_j^2 , ($j = 1, \dots, D$), the integral is given by

$$\begin{aligned} I &= \int_{\Omega} f(\vec{\mu}) d\vec{\mu} \\ &= \int_{\Omega} \left[\prod_{i=1}^m \frac{1}{(2\pi)^{D/2} |\Sigma|^{1/2}} e^{-\frac{1}{2}(\vec{x}_i - \vec{\mu})^T \Sigma^{-1} (\vec{x}_i - \vec{\mu})} \right] d\vec{\mu} \\ &= \frac{1}{(2\pi)^{D \cdot (m-1)/2} |\Sigma|^{(m-1)/2} m^{D/2}} \exp \left(- \sum_{j=1}^D \frac{\text{Var}[x_j]}{2\sigma_j^2} \right) \end{aligned}$$

where Σ is the (diagonal) covariance matrix.

For our concrete example, we take $m = 100$ and generate random values of x from a Gauss distribution of mean zero and unit standard deviation, and we find for the generated values $\ln I_{\text{true}} = -67.18$. In evaluating the integral, we take for the support $\Omega = \{-100 \leq \mu \leq 100\}$.

We then use $N_{\text{MCMC}} = 10^5$ samples from the MCMC output to find an estimate for the mode of $f(\mu)$ and to calculate the standard deviation for μ . The distribution of samples from $f(\mu)$ from the MCMC are displayed in Fig. 1(left). The mode of the samples is found at $\mu^* = -0.05$ and the standard deviation is found to be $\sigma_{\mu} = 0.10$. The effective sample size for this set of samples is $N_{\text{ESS}} = 3.9 \cdot 10^4$.

The dependence of \hat{r} on the chosen value of Δ is also shown in Fig. 1(right) for 500 values of Δ ranging from 0.01 to 5 in steps of 0.01. For a one-dimensional Gaussian target density, which is what we have here, the expectation is that 68 % of MCMC samples occur within $\Delta = 1$ and 95 % occur within $\Delta = 2$, and this is indeed what is found.

We then perform a sample mean calculation with $N_{\text{SM}} = 10^5$ samples for each of the different choices of Δ . For each calculation, we extract a value of \hat{I}_{AME} as described in section 2.3 as well as an estimate of the uncertainty. The extracted values of \hat{I}_{AME} (divided by the true value) are shown as a function of Δ in Fig. 2(left). The error bars are the estimated one standard deviation uncertainties. We observe small systematic deviations of the results for small values of Δ resulting from the inaccurate determination of r from the MCMC samples (note that the MCMC was only run once, so that the \hat{r} values are correlated).

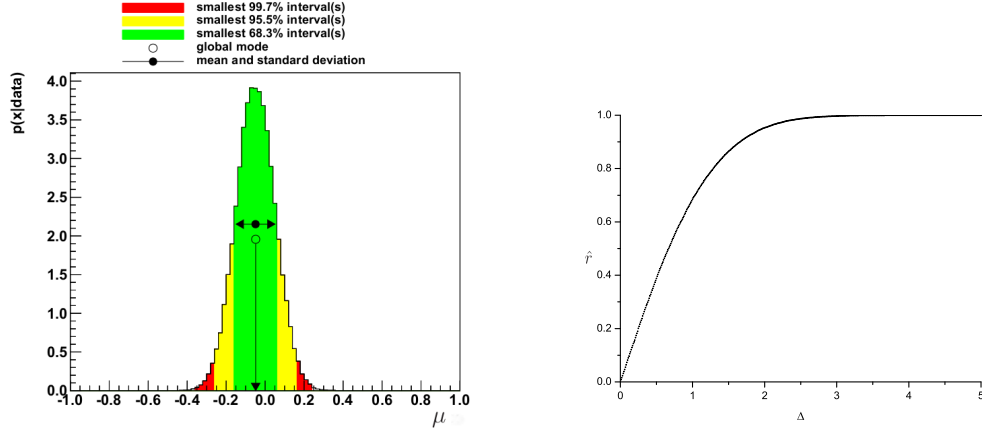


FIGURE 1. Left) Distribution of samples from the MCMC algorithm BAT [13] for the product of 100 Gauss distributions as described in the text. Right) Fraction of MCMC samples falling within the interval of length 2Δ as a function of the value of Δ (in units of the standard deviation of the distribution).

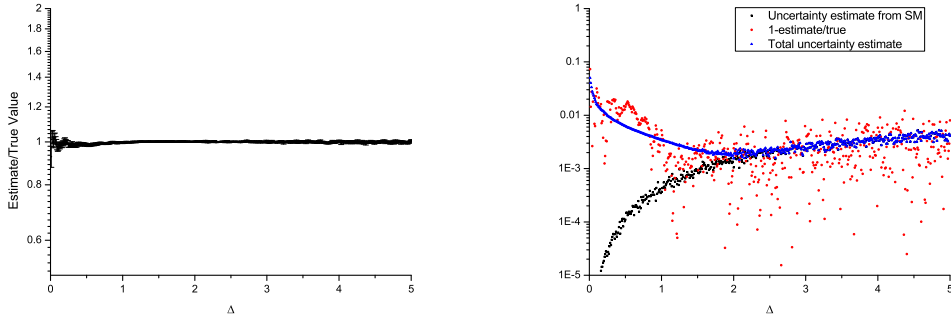


FIGURE 2. Arithmetic Mean results for the one-dimensional Gauss product example. Left) \hat{I}_{AME} as a function of Δ , scaled by the true value. The error bars correspond to the estimated uncertainty. Right) The actual error $|\hat{I}_{\text{AME}}/I_{\text{true}} - 1|$ (red), the estimated uncertainty $\sigma_{I_{\Delta}}/I_{\text{true}}$ (black) and the total estimated uncertainty σ_I/I_{true} (blue) as a function of Δ .

To study the uncertainty estimation, we compare σ_I/I_{true} to $(\hat{I}_{\text{AME}} - I_{\text{true}})/I_{\text{true}}$ at each value of Δ . The results are shown in Fig. 2(right). In this figure, the red points indicate the absolute value of $(\hat{I}_{\text{AME}} - I_{\text{true}})/I_{\text{true}}$, the black points the estimated uncertainty coming from the sample mean calculation, $\sigma_{I_{\Delta}}/I_{\text{true}}$, and the blue points the total estimated uncertainty, σ_I/I_{true} . We observe that our estimated uncertainty is accurate, and that there is a

minimum of the uncertainty around $\Delta = 2$. The location of the minimum clearly depends on the number of samples chosen for the MCMC and sample mean calculations, but it is important that we can accurately estimate the uncertainty. In this case, the arithmetic mean calculation is quite accurate even at large values of Δ since we are only working in one-dimension.

We now evaluate the Harmonic Mean estimate for I as described in section 2.5. The estimate \hat{I}_{HME} as well as the absolute deviation from I_{true} as a function of Δ are shown in Fig. 3. We see for this example that the HME technique works well, and that accuracies of a fraction of 1 % are possible from the HME estimation at $\Delta \approx 1.5$. As Δ is increased, the HME estimation worsens since, although more of the MCMC samples are included, reducing the binomial uncertainty on N_{Δ} , imperfect sampling in the tails of the distribution plays a large role and we see the importance of limiting the range of the integration region for the HME calculation already with this simple one-dimensional example. The uncertainty is somewhat worse than what was found for the AME calculation, but probably adequate for the majority of applications. Also, the calculation did not require the extra step of performing a sample mean calculation.

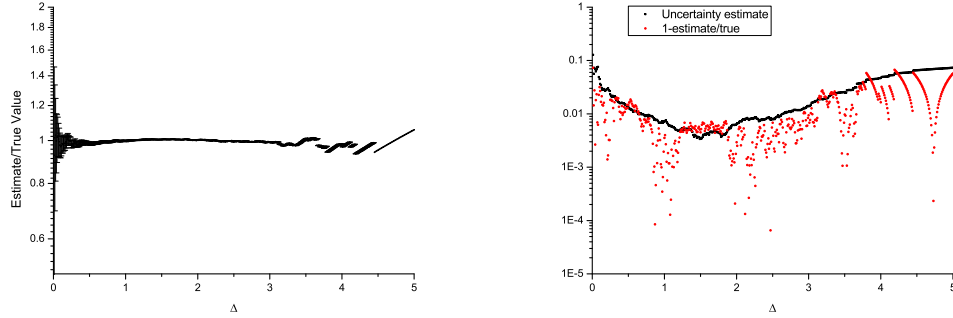


FIGURE 3. Harmonic Mean results for the one-dimensional Gauss product example. Left) \hat{I}_{HME} scaled by the true value as a function of Δ . The error bars correspond to the estimated uncertainty. Right) The actual error $|\hat{I}_{\text{HME}}/I_{\text{true}} - 1|$ and the estimated uncertainty as a function of Δ .

As seen in Fig. 1, the target density is Gaussian and therefore the Laplace method is expected to work well. Indeed, the Laplace method yields an estimate within 0.2 % of the true value in this example: $\hat{I}_{\text{L}}/I_{\text{true}} = 1.0016$.

3.2. Product of Multivariate Gaussians. We now move to a target density composed of a product of ten dimensional Gaussian distributions with non-diagonal covariance matrix. The target function in this case is:

$$(19) \quad f(\vec{\mu}) = \prod_{i=1}^m \frac{1}{(2\pi)^5 |\Sigma|^{1/2}} e^{-\frac{1}{2}(\vec{x}_i - \vec{\mu})^T \Sigma^{-1}(\vec{x}_i - \vec{\mu})}$$

where Σ is the covariance matrix, assumed to be known, and $m = 20$. The target function is ten-dimensional and has significant correlations among the ten parameters. The values of \vec{x}_i were chosen by generating random vectors using $\vec{\mu}_{\text{true}} = \vec{0}$ and the following covariance matrix

$$\Sigma = \begin{bmatrix} 1.0 & 1.0 & 0.30 & 0.43 & -0.14 & -0.86 & -0.22 & -0.84 & 0.83 & -2.5 \\ 1.0 & 2.0 & -0.14 & 0.36 & 0.14 & -0.08 & -0.45 & 0.71 & 0.07 & -1.9 \\ 0.30 & -0.14 & 3.0 & -0.04 & 0.76 & 0.85 & 1.7 & -0.41 & 1.2 & -0.78 \\ 0.43 & 0.36 & -0.04 & 4.0 & -0.81 & -0.86 & -1.5 & -2.1 & 0.17 & 0.63 \\ -0.14 & 0.14 & 0.76 & -0.81 & 5.0 & 2.5 & 1.6 & 1.5 & 1.9 & 0.46 \\ -0.86 & -0.08 & 0.85 & -0.86 & 2.5 & 6.0 & 1.9 & 4.1 & -0.28 & 2.7 \\ -0.22 & -0.45 & 1.7 & -1.5 & 1.6 & 1.9 & 7.0 & 0.70 & 1.4 & 2.6 \\ -0.84 & 0.71 & -0.41 & -2.1 & 1.5 & 4.1 & 0.70 & 8.0 & -0.87 & 2.5 \\ 0.83 & 0.07 & 1.2 & 0.17 & 1.9 & -0.28 & 1.4 & -0.87 & 9.0 & -4.0 \\ -2.5 & -1.9 & -0.78 & 0.63 & 0.46 & 2.7 & 2.6 & 2.5 & -4.0 & 10 \end{bmatrix}$$

and again could represent a type of situation found in a data analysis setting. The integration region for I was taken as a 10D hypercube of side length 200 centered on $\vec{\mu} = \vec{0}$.

The value for I_{true} can again be evaluated analytically by finding the similarity transformation that diagonalizes the covariance matrix. The expression of the integral in this case is

$$(20) \quad I = \frac{A}{(2\pi)^{N \cdot (m-1)/2} |\Sigma|^{(m-1)/2} m^{N/2}} \exp \left(- \sum_{j=1}^N \text{Var}[x'_j] \right).$$

where $\vec{x}' = \sqrt{S}V\vec{x}$ and $\Sigma^{-1} = V^{-1}SV$ with S a diagonal matrix.

The true value of the integral for randomly generated data was evaluated using this expression and yielded $\ln(I_{\text{true}}) = -427.5$. The MCMC program BAT was used to sample from the target density with $5 \cdot 10^5$ samples stored post-convergence (yielding $N_{\text{ESS}} = 1.9 \cdot 10^5$). The value of \hat{r} is given as a function of Δ in Fig. 4.

The arithmetic mean calculation was performed at each of 500 values of Δ as in the one-dimensional case, with 10^5 samples in each AME run. The results are shown in Fig. 5. As is seen, for values of Δ around 1.5, the uncertainty is about 1 %. The method does not show any systematic biases for $\Delta > 1$, and the estimated uncertainty is again a good estimator for the error. At small Δ , where a small number of MCMC samples are used, the correlation between the MCMC samples produces some systematic errors in the evaluation of I .

The results for the HME estimator are also shown in Fig. 5. We see that accuracies of a few tens of % are achieved, but only in a narrow Δ range. For $\Delta > 2$, the error is more than 100 % and the HME estimate is no longer useful. Also, the estimated uncertainty is too low and does not provide a reliable estimate of the true error. The HME method is clearly already running into trouble at this level of complexity.

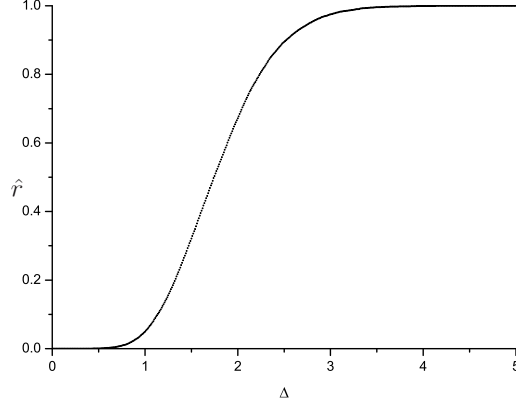


FIGURE 4. Fraction of MCMC samples falling within the hypercube centered on the observed mode and of side length 2Δ as a function of the value of Δ (in units of the standard deviation of the marginalized distribution) for the product of ten-dimensional correlated Gauss functions.

The target density is again a multivariate Gaussian, and the Laplace method works well, yielding $\hat{I}_L/I_{\text{true}} = 0.977$.

3.3. Gaussian Shell. We now move beyond simple unimodal Gaussian type target densities and consider a function in D dimensions with degenerate modes lying on a $D - 1$ dimensional surface of fixed radius, a Gaussian shell:

$$(21) \quad f(\vec{\lambda}|\vec{c}, r, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(|\vec{\lambda} - \vec{c}| - r)^2}{2\sigma^2}\right).$$

This function is centered at \vec{c} with degenerate modes along a surface of radius r . The value of the function decreases away from the modal surface along a radius according to a Gaussian shape with standard deviation σ . The integral of this function can be evaluated using spherical coordinates centered at \vec{c} , where ρ is the radial coordinate in the space, so that

$$I = \frac{1}{\sqrt{2\pi\sigma^2}} \int \exp\left(-\frac{(\rho - r)^2}{2\sigma^2}\right) dV$$

The volume element, integrated over the angular coordinates, is $dV = S_{D-1}\rho^{D-1}d\rho$ with $S_{D-1} = 2\pi^{D/2}/\Gamma(D/2)$, so that we have

$$I = \frac{\sqrt{2}\pi^{(D-1)/2}}{\Gamma(D/2)\sigma} \int_0^{\rho_{max}} \rho^{D-1} \exp\left(-\frac{(\rho - r)^2}{2\sigma^2}\right) d\rho.$$

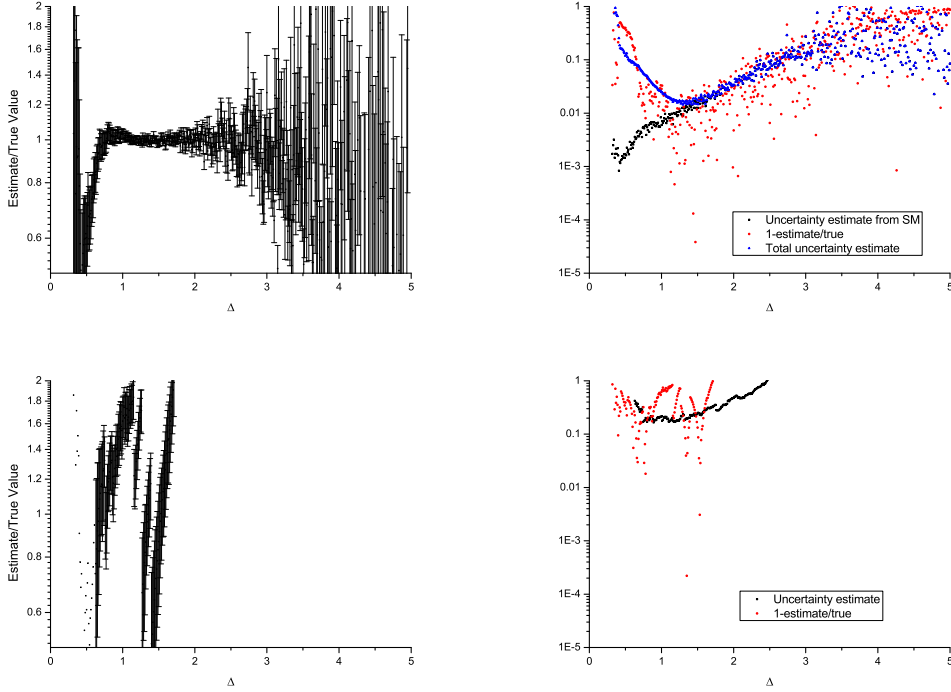


FIGURE 5. 10D correlated Gauss product example. Top left) \hat{I}_{AME} as a function of Δ , scaled by the true value. Top right) The actual error $|\hat{I}_{\text{AME}}/I_{\text{true}} - 1|$ (red), the estimated uncertainty from the sample mean calculation (black) and the total estimated uncertainty (blue) as a function of Δ . Bottom left) \hat{I}_{HME} scaled by the true value as a function of Δ . Bottom right) The actual error $|\hat{I}_{\text{HME}}/I_{\text{true}} - 1|$ and the estimated uncertainty as a function of Δ . The error bars in the left plots correspond to the estimated uncertainty.

We are left with a one-dimensional integral that can be easily calculated numerically to high precision. Note that we have assumed that the integral in the region outside ρ_{max} (the corners in the hypercube) is vanishingly small. This is the case for the examples considered in this article.

For the three examples below, we use the following settings: radius $r = 5$, width $\sigma = 2$ and $\vec{c} = \vec{0}$. The integration region extends from $-10, +10$ in each dimension.

3.3.1. 2-Dimensional Gaussian-shell. The parameter values result in $I_{\text{true}} = 7.86 \cdot 10^{-4}$. We use the BAT code to produce 10^5 MCMC samples from the target density, yielding an effective sample size $N_{\text{ESS}} = 3.9 \cdot 10^4$. The sample distribution from the MCMC as well as the estimate of \hat{r} as a function of Δ are shown in Fig. 6.

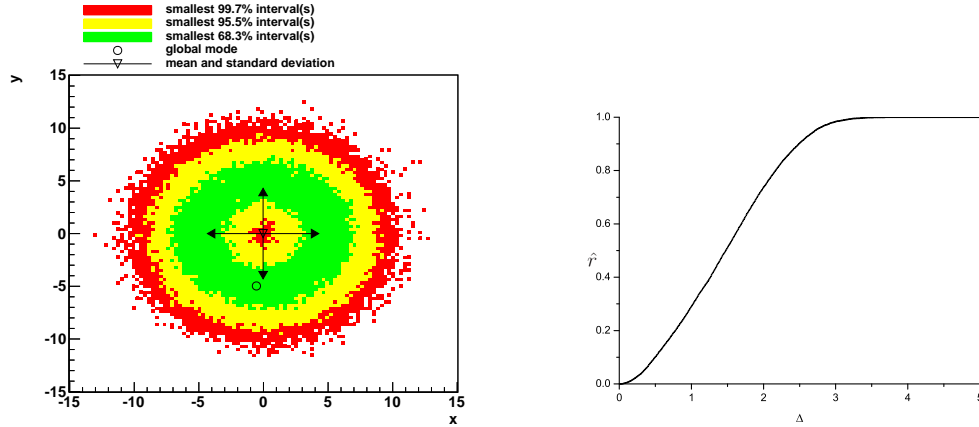


FIGURE 6. Left) Distribution of samples from the MCMC algorithm BAT [13] for the two-dimensional shell example. The small circle indicates the location of the mode found from the MCMC samples. The arrows indicate the mean and ± 1 standard deviation ranges. Right) Fraction of MCMC samples falling within the hypercube of side length 2Δ as a function of the value of Δ (in units of the standard deviation of the distribution).

As can be seen in the figure, the MCMC has produced a reasonable sample distribution. The location of the mode from the posterior samples happens to be close to $(0, -5)$ and is indicated in the figure (note that $\vec{\lambda} = (x, y)$ in the figure). The lack of a single mode is not a problem for the AME and HME algorithms, but we no longer expect the Laplace method to give sensible results. The mean values of (x, y) are very close to $(0, 0)$ and the standard deviation in each direction is about 5 units. The hypercube centered at the mode found from the MCMC samples and with $\Delta = 1$ contains about 20 % of the samples, and the hypercube with $\Delta = 2$ contains about 75 % of the samples.

We again use 10^5 samples for our sample mean calculations at each of the values of Δ . The results for \hat{I}_{AME} are shown in the top plots in Fig. 7, and we see that there is no difficulty in achieving a good result for the integral despite not having a simple mode for the target distribution. The accuracy of the calculation is good, and the uncertainty is better than 1 % for a wide range of Δ , despite the rather small number of samples in the MCMC and AME calculations. We again find that our estimated uncertainty gives a good reproduction of the actual error.

The HME evaluations are also given in Figs. 7. Here we find good performance (few % level accuracy) up to $\Delta = 2$, at which point the HME calculation starts to systematically deviate from the correct value. In this case, the estimated uncertainty does not give a reliable indication of the actual error for $\Delta > 2$ and in fact the uncertainty is grossly underestimated. This is a result of the missing MCMC samples at very small $f(\vec{\lambda})$. The volume term in the numerator in Eq. 16 grows as Δ is increased, but is not properly

compensated by large terms that should appear in the denominator from small values of $f(\vec{\lambda})$. The inability to diagnose this behavior implies that the HME is unreliable.

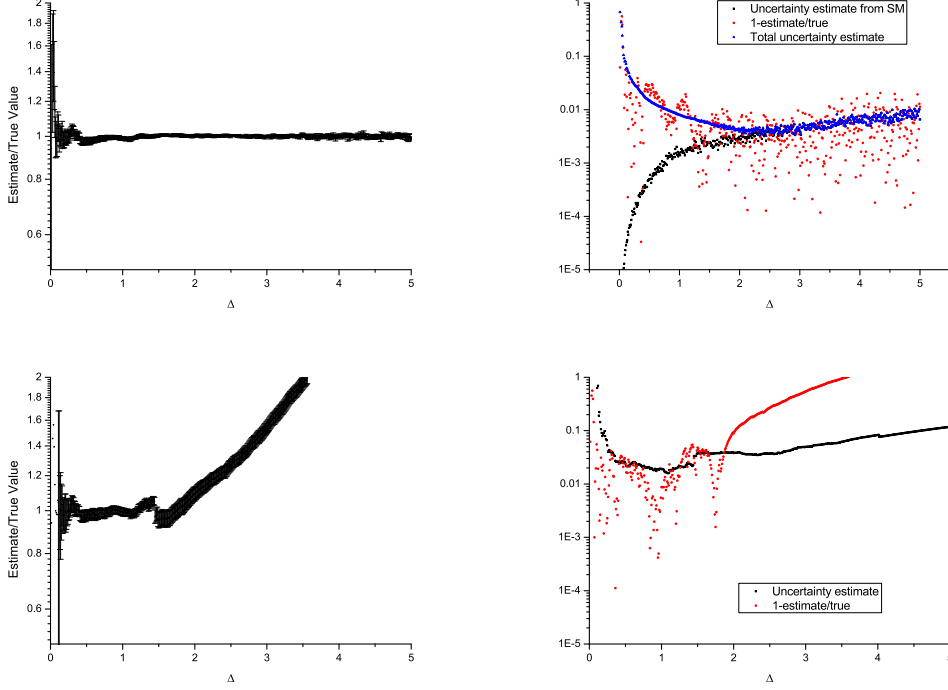


FIGURE 7. Two-dimensional Gaussian shell example. Top left) \hat{I}_{AME} as a function of Δ , scaled by the true value. Top right) The actual error $|\hat{I}_{\text{AME}}/I_{\text{true}} - 1|$ (red), the estimated uncertainty from the sample mean calculation (black) and the total estimated uncertainty (blue) as a function of Δ . Bottom left) \hat{I}_{HME} scaled by the true value as a function of Δ . Bottom right) The actual error $|\hat{I}_{\text{HME}}/I_{\text{true}} - 1|$ and the estimated uncertainty as a function of Δ . The error bars in the left plots correspond to the estimated uncertainty.

As expected, the Laplace method does not work for the Gaussian shell situation. For the two-dimensional example considered here, $\hat{I}_{\text{L}}/I_{\text{true}} = 0.03$.

3.3.2. 10-Dimensional Gaussian-shell. Here $\ln(I_{\text{true}}) = -32.16$. In a first calculation, we use the BAT code to produce 10^5 MCMC samples from the target density, yielding an effective sample size $N_{\text{ESS}} = 3.9 \cdot 10^4$ and calculate the evidence. We again use 10^5 samples for our sample mean calculations. The results for the AME and HME evaluations are given in Fig. 8. For the arithmetic mean calculation, we see the same pattern as in the previous examples. For small values of Δ , the uncertainty coming from the small number of MCMC

samples dominates. However, sub % errors are possible for $\Delta \sim 1.5$, which corresponds to $\hat{r} \approx 0.2$. As Δ increases, the uncertainties from the sample mean calculation dominate since we move to regions of the space that do not contain significant probability mass. The estimated uncertainty is again accurate and can be used as a guide to choose the optimal value of Δ as we discuss below.

The HME estimate achieves few % accuracy at a somewhat smaller value of Δ than the optimal for the sample mean calculation. The estimated uncertainty is again tends too small at larger Δ and is not reliable.

As expected, the Laplace method does not work well and yields $\hat{I}_L/I_{\text{true}} = 0.08$.

As a check that these results are not due to small MCMC sample size, the calculations were redone for 10^6 MCMC samples. The optimal value of Δ changes somewhat for the sample mean calculation, but otherwise all results are basically as before. The systematic behavior of the $|\hat{I}_{\text{HME}}/I_{\text{true}} - 1|$ is the same as for the smaller MCMC sample size; no significant improvement in performance was found with the 10 times large MCMC sample size.

3.3.3. 50-Dimensional Gaussian-shell. As an extreme example, we considered a 50-dimensional Gaussian shell. Here the modal surface is a 49-dimensional hypersphere and $\ln(I_{\text{true}}) = -169.8$. The BAT code was used to initially produce 10^5 MCMC samples from the target density, yielding an effective sample size $N_{\text{ESS}} = 3.9 \cdot 10^4$. The values of \hat{r} increase rapidly from $\hat{r} \approx 0$ at $\Delta = 1.8$ to $\hat{r} \approx 0.85$ at $\Delta = 3$. The standard deviations in each dimension is about 2 units, so that $\Delta = 5$ approximately covers the full support defined for the function. The results for the AME and HME evaluations are given in Fig. 9. The best result for the sample mean calculation gives about 10 % accuracy, whereas the HME calculation is within 100 % of the correct result for a small range of Δ where \hat{r} starts to increase.

We used 10^5 samples for our sample mean calculations, although this is clearly too small a number for such a large dimensional volume. The error from the sample mean calculation increases rapidly as we increase Δ , and becomes completely unreliable for $\Delta > 2$. For such a large volume, the vast majority of sample mean evaluations are in regions where the target density is vanishingly small and the uncertainty grossly underestimates the true error. In the next section, we discuss a choice of settings for the sample mean calculation and redo the calculation shown here.

As expected, the Laplace method does not work well and yields $\hat{I}_L/I_{\text{true}} = 6 \cdot 10^{-4}$.

We again checked that these results are not due to small MCMC sample size, the calculations were redone for 10^7 MCMC samples. The optimal location of Δ changes to smaller values for the sample mean calculation and few % level accuracy is reached. For the HME calculation, a small improvement is also observed, but otherwise all results are basically as before.

4. DISCUSSION

Based on the results in the previous sections, we discuss now a procedure for choosing the value of Δ for both the sample mean and harmonic mean estimators. As was seen in our examples, the uncertainty in the calculation for the AME estimator comes from

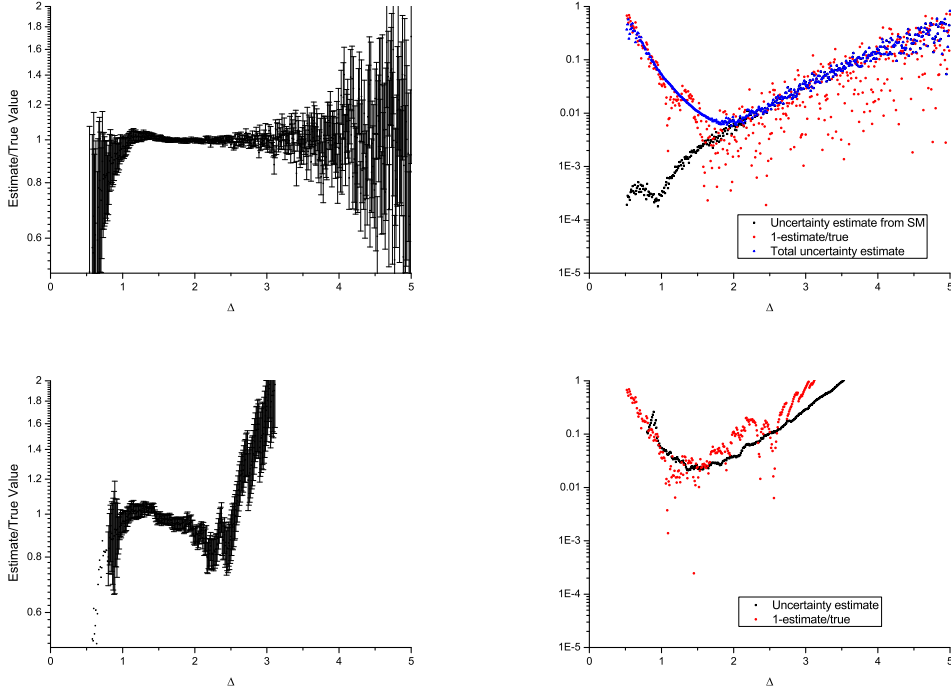


FIGURE 8. Ten-dimensional Gaussian shell example. Top left) \hat{I}_{AME} as a function of Δ , scaled by the true value. Top right) The actual error $|\hat{I}_{\text{AME}}/I_{\text{true}} - 1|$ (red), the estimated uncertainty from the sample mean calculation (black) and the total estimated uncertainty (blue) as a function of Δ . Bottom left) \hat{I}_{HME} scaled by the true value as a function of Δ . Bottom right) The actual error $|\hat{I}_{\text{HME}}/I_{\text{true}} - 1|$ and the estimated uncertainty as a function of Δ . The error bars in the left plots correspond to the estimated uncertainty.

two sources - the approximately binomial fluctuations in the number of MCMC samples included in our region of interest specified by Δ , and the uncertainty coming from the sample mean calculation. The first uncertainty can be estimated from the MCMC output, and can be used to define a value of \hat{r} by specifying that this source of uncertainty should contribute half of the final uncertainty. I.e., we find the value of \hat{r} such that (see Eq. 7)

$$\frac{\sigma_r}{\hat{r}} = \frac{\epsilon}{\sqrt{2}}$$

where ϵ is the target uncertainty. We will use $\epsilon = 0.01$ for our discussion below except for the fifty-dimensional Gaussian shell example, where we take $\epsilon = 0.05$. Once we have fixed \hat{r} in this way, we then find the corresponding value of Δ and use this to calculate sample

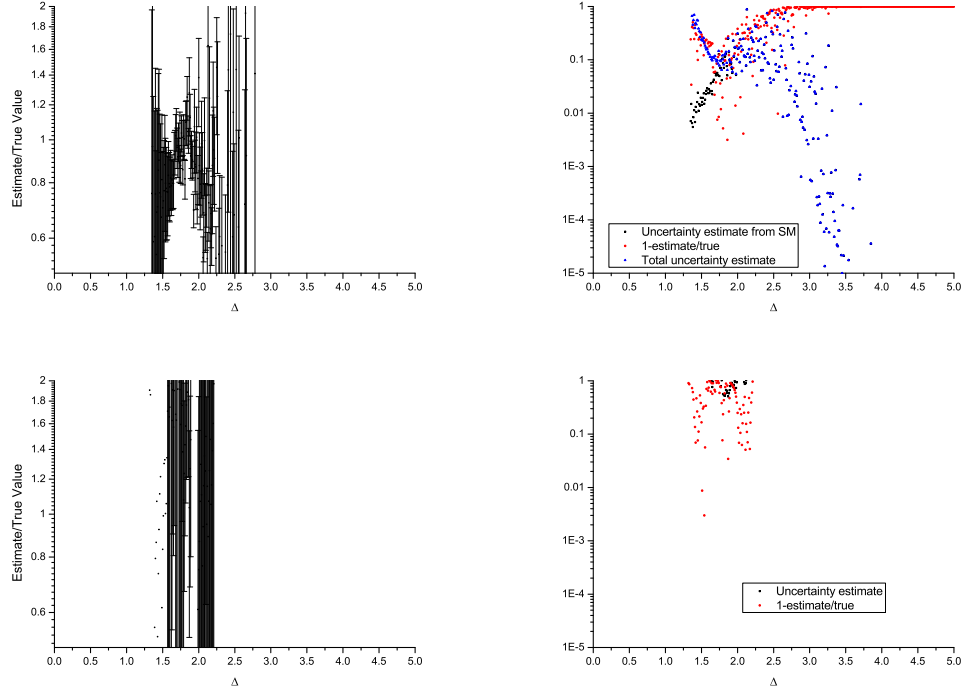


FIGURE 9. 50-dimensional Gaussian shell example. Top left) \hat{I}_{AME} as a function of Δ , scaled by the true value. Top right) The actual error $|\hat{I}_{\text{AME}}/I_{\text{true}} - 1|$ (red), the estimated uncertainty from the sample mean calculation (black) and the total estimated uncertainty (blue) as a function of Δ . Bottom left) \hat{I}_{HME} scaled by the true value as a function of Δ . Bottom right) The actual error $|\hat{I}_{\text{HME}}/I_{\text{true}} - 1|$ and the estimated uncertainty as a function of Δ . The error bars in the left plots correspond to the estimated uncertainty.

mean integrals with for a batch of samples, requiring a minimum of 10 batches. We use the variance of these 10 calculations to determine how many batches will be needed to get the desired uncertainty; i.e.,

$$\frac{\sigma_{I_{\Delta}}}{\hat{I}_{\Delta}} = \frac{\epsilon}{\sqrt{2}}.$$

The results for the examples given in the previous sections using this procedure for fixing the parameters of the algorithm is given in Table 1. As is seen, the range of values for Δ is relatively narrow and only grows slowly with the complexity of the target function. The number of sample mean calculations however depends strongly on the complexity of the problem, and is also inversely dependent on the accuracy specified and on the size of the MCMC sample. For a given specified accuracy, Δ is reduced as N_{MCMC} is increased,

| Test | N_{MCMC} | N_{ESS} | ϵ | Δ | N_{SM} | I_{true} | $1 - \frac{\hat{I}_{\text{AME}}}{I_{\text{true}}}$ | $\sigma_{I_{\text{AME}}}/I_{\text{AME}}$ |
|--------------|-------------------|------------------|------------|----------|-----------------|------------------------|--|--|
| 1D Gaussian | $1 \cdot 10^5$ | $3.9 \cdot 10^4$ | 0.01 | 0.44 | 18 | $6.46 \cdot 10^{-68}$ | 0.011 | 0.010 |
| 10D Gaussian | $1 \cdot 10^5$ | $3.5 \cdot 10^4$ | 0.01 | 1.56 | 652000 | $9.09 \cdot 10^{-124}$ | 0.008 | 0.010 |
| 2D shell | $1 \cdot 10^5$ | $3.9 \cdot 10^4$ | 0.01 | 1.12 | 6510 | $7.71 \cdot 10^{-4}$ | 0.009 | 0.010 |
| 10D shell | $1 \cdot 10^5$ | $3.9 \cdot 10^4$ | 0.01 | 1.69 | 18700 | $1.085 \cdot 10^{-14}$ | 0.009 | 0.010 |
| 50D shell | $1 \cdot 10^5$ | $3.9 \cdot 10^4$ | 0.05 | 1.89 | 1352400 | $1.81 \cdot 10^{-74}$ | 0.056 | 0.046 |
| 10D shell | $1 \cdot 10^6$ | $8.9 \cdot 10^5$ | 0.01 | 1.13 | 718 | $1.085 \cdot 10^{-14}$ | 0.010 | 0.010 |
| 50D shell | $1 \cdot 10^7$ | $8.6 \cdot 10^6$ | 0.05 | 1.47 | 19200 | $1.81 \cdot 10^{-74}$ | 0.090 | 0.044 |

TABLE 1. Summary of the results on different target functions for the AME estimator of the normalizing integral. N_{MCMC} is the number of posterior samples from the MCMC, N_{ESS} is the effective sample size, ϵ is the specified accuracy for the integral calculation, Δ is the multiplier of the standard deviation along each dimension chosen by the algorithm, N_{SM} is the number of samplings of the function used in the sample mean calculation, I_{true} is the true value of the integral, $1 - \frac{\hat{I}_{\text{AME}}}{I_{\text{true}}}$ is the fractional error made in the calculation and $\sigma_{I_{\text{AME}}}/I_{\text{AME}}$ is the estimated fractional uncertainty from the calculation.

| Test | N_{MCMC} | N_{ESS} | Δ | I_{true} | $1 - \frac{\hat{I}_{\text{HME}}}{I_{\text{true}}}$ | $\sigma_{I_{\text{HME}}}/I_{\text{HME}}$ |
|--------------|-------------------|------------------|----------|------------------------|--|--|
| 1D Gaussian | $1 \cdot 10^5$ | $3.9 \cdot 10^4$ | 0.66 | $6.46 \cdot 10^{-68}$ | 0.008 | 0.010 |
| 10D Gaussian | $1 \cdot 10^5$ | $3.5 \cdot 10^4$ | 1.74 | $9.09 \cdot 10^{-124}$ | 1.129 | 0.353 |
| 2D shell | $1 \cdot 10^5$ | $3.9 \cdot 10^4$ | 1.47 | $7.71 \cdot 10^{-4}$ | 0.019 | 0.034 |
| 10D shell | $1 \cdot 10^5$ | $3.9 \cdot 10^4$ | 1.85 | $1.085 \cdot 10^{-14}$ | 0.066 | 0.031 |
| 50D shell | $1 \cdot 10^5$ | $3.9 \cdot 10^4$ | 2.46 | $1.81 \cdot 10^{-74}$ | 23.251 | 17.482 |
| 10D shell | $1 \cdot 10^6$ | $8.9 \cdot 10^5$ | 1.85 | $1.085 \cdot 10^{-14}$ | 0.005 | 0.004 |
| 50D shell | $1 \cdot 10^7$ | $8.6 \cdot 10^6$ | 2.46 | $1.81 \cdot 10^{-74}$ | 2.680 | 0.707 |

TABLE 2. Summary of the results on different target functions for the HME estimator of the normalizing integral with \hat{r} set to 0.5. N_{MCMC} is the number of posterior samples from the MCMC, N_{ESS} is the effective sample size, Δ is the multiplier of the standard deviation along each dimension chosen by the algorithm, I_{true} is the true value of the integral, $1 - \frac{\hat{I}_{\text{HME}}}{I_{\text{true}}}$ is the fractional error made in the calculation and $\sigma_{I_{\text{HME}}}/I_{\text{HME}}$ is the estimated fractional uncertainty from the calculation.

and this reduces the number of sample mean calculations necessary. We find that the AME algorithm gives a reliable estimate of the uncertainty for the examples chosen if the required number of sample mean calculations is not too large. We conclude that the AME calculation of the integral of the target density using a reduced volume around the mode of the target works well for the types of cases we have studied.

For the HME estimator, the post-convergence samples of the MCMC are used in the calculation. For want of a better method, we fix Δ by requiring that $\hat{r} = 0.5$ as this value was typically near optimal for the examples studied. The results for the examples given above using this fixing of the algorithm is given in Table 2.

As can be seen from the table, and as discussed earlier, the HME calculation works well for the simple target functions considered, but does not produce good results for the more complicated target functions. In particular, the estimated uncertainty does not provide a good estimate of the actual error, so that it is not possible to diagnose that the calculation is not performing well. We therefore do not recommend the use of the HME estimator to calculate the normalization integral for anything but the simplest low-dimensional target densities,

The Laplace estimation works well in cases where the target density is well approximated by a (multivariate) Gaussian distribution. If this is known to be the case, then this approximation is easily calculated and can be used. However, it should be avoided if the shape of the target distribution is not well known.

5. SUMMARY

We have investigated techniques for the integration of the target density in cases where a MCMC algorithm has successfully run. We do not attempt to modify the sampling of the target density, but only to provide a post-processor for an MCMC algorithm. From the MCMC, we have an estimate of the global mode and also the variance of the samples marginalized along each parameter dimension. We use this information to define a hypercube centered on the global mode and having side lengths proportional to the standard deviation along these directions, and then calculate the integral of the target function in the reduced volume using either an arithmetic mean or harmonic mean approach. The fraction of MCMC samples within the reduced volume was used to estimate the integral of the target density over the full volume of interest. This technique was tried on a variety of examples and also compared to a Laplace estimator. The key elements of the methods studied are:

- Given the MCMC has been run successfully, the evaluation of the normalization of the target function can be performed using any sub support of the support of the target function;
- From the MCMC, we can find a point near the maximum of the target function, and we can perform the integration in a region which is in some ways optimal by centering the sub support on this point;
- It is possible to also calculate an estimated accuracy for the integral.

Our conclusions are that the arithmetic mean calculation performed in a hypercube centered on the observed mode works well and provides a technique for calculating the normalization of the target density with a reliable uncertainty estimate. On the other hand, the harmonic mean estimator only works well in situations where the range of values from the target density does not vary too widely, and the Laplace estimator is restricted for use on Gaussian shaped target distributions.

ACKNOWLEDGMENTS

The authors would like to thank Frederik Beaujean, Daniel Greenwald, Stephan Jahn and Kevin Kröninger for many fruitful discussions.

REFERENCES

- [1] See e.g., C. Robert and G. Casella, ‘Monte Carlo Statistical Methods’, 2nd Edition, Springer (2004).
- [2] H. Jeffreys, ‘Theory of Probability’, 3rd Ed., Clarendon Press, Oxford, MR0187257 (1961).
- [3] E. T. Jaynes, ‘Probability Theory: the logic of science’, Cambridge University Press, Cambridge, MR1992316 (2003).
- [4] D. N. VanDerwerken, S. C. Schmidler, ‘Parallel Markov Chain Monte Carlo’, arXiv:1312.7479v1
- [5] N. Friel and J. Wyse, ‘Estimating the evidence - a review’, Stat. Neerl., **66** (2012) 2800.
- [6] L. Tierney and J. B. Kadane, ‘Accurate approximations for posterior moments and marginal densities’, Journal of the American Statistical Associations, **81** (1986) 82.
- [7] M. A. Newton and A. E. Raftery, ‘Approximate Bayesian Inference with the Weighted Likelihood Bootstrap’, Journal of the Royal Statistical Society, series B**56** (1994) 3.
- [8] S. Chib, I. Jeliazkov, ‘Marginal Likelihood from the Metropolis-Hastings Output’, Journal of the American Statistical Association **96** (2001) 270.
- [9] C. P. Robert and D. Wraith, ‘Computational Methods for Bayesian model choice’, Bayesian Inference and Maximum Entropy Methods in Science and Engineering: the 29th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering (AIP Conference Proceedings), vol. 1193 (2009) 251.
- [10] J. Skilling, ‘Nested Sampling for General Bayesian Computation’, Bayesian Analysis **1** (2006) 833.
- [11] A. Gelman and X. L. Meng, ‘Simulating normalizing constants: from importance sampling to bridge sampling to path sampling’, Statistical Science **13** (1998) 163.
- [12] N. Friel and A. N. Pettitt, ‘Marginal likelihood estimation via power posteriors’, Journal of the Royal Statistical Society, Series B**70** (2008) 589.
- [13] A. Caldwell, D. Kollar, K. Kröninger, ‘BAT - The Bayesian Analysis Toolkit’, Comput. Phys. Commun. **180** (2009) 2197-2209.
- [14] R. E. Kass, B. P. Carlin, A. Gelman, and R. Neal, ‘Markov chain Monte Carlo in practice: A roundtable discussion’, The American Statistician, **52** (1998) 93.
- [15] A. E. Gelfand and D. K. Dey, ‘Bayesian model choice: asymptotics and exact calculations’, Journal of the Royal Statistical Society, B **56** (1994) 501-514.